Quantifying Uncertainty In Computational Knowledge Engineering Rapidly

Adam Donato, Professor Ranga Pitchumani, and Dr. Mehrdad Shahnam

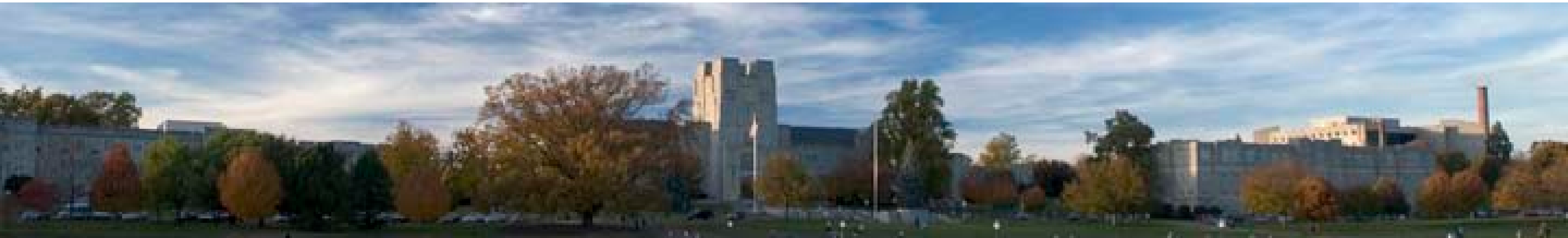
Advanced Materials and Technologies Laboratory

Department of Mechanical Engineering

Virginia Tech

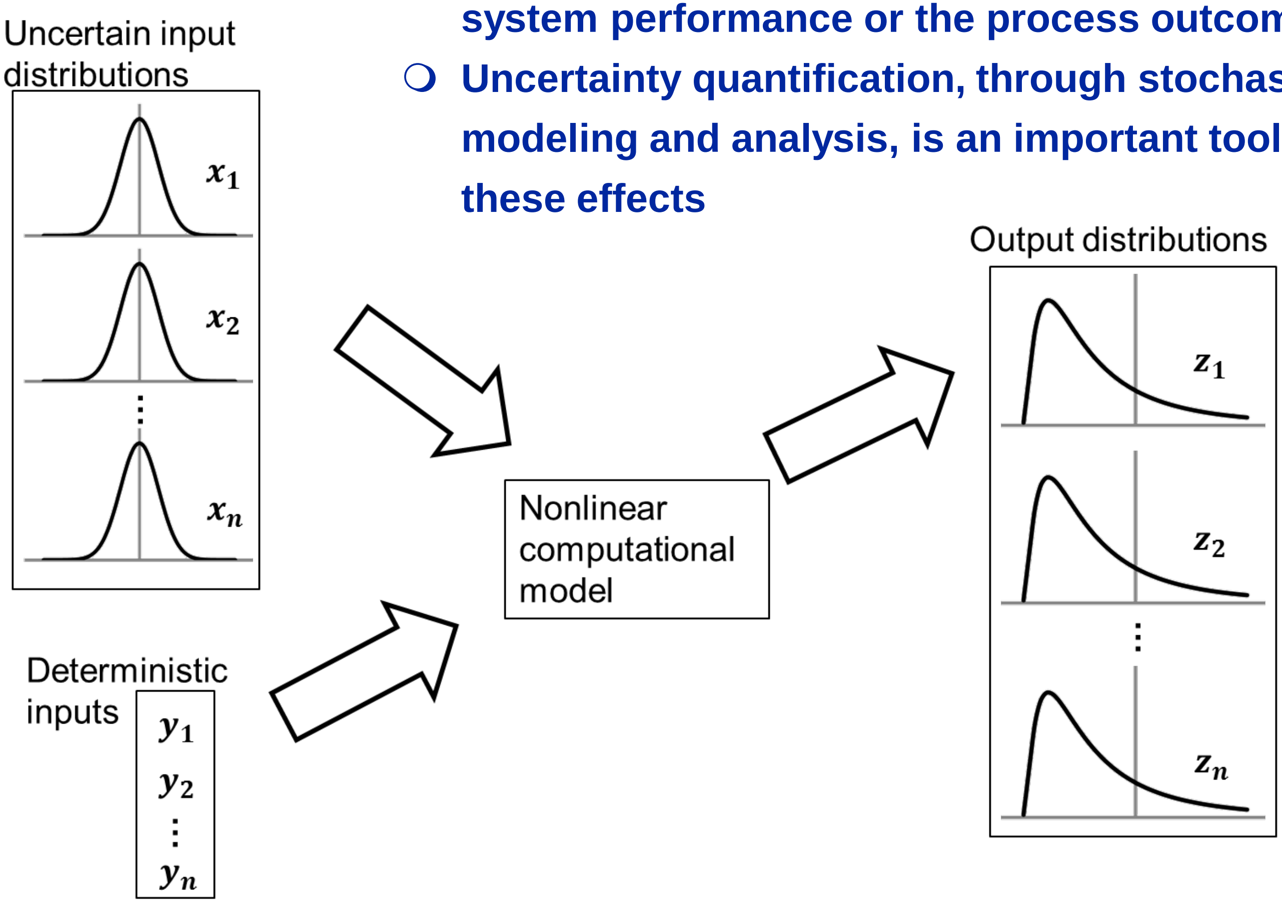
Blacksburg, Virginia 24061-0238

adonato@vt.edu • pitchu@vt.edu • http://www.me.vt.edu/amtl



Motivation for Uncertainty Quantification

- **Multiphase systems operate in an environment of uncertainty**
  - **This uncertainty exists in both the parameters governing the system and in the process behavior**
  - **The interactive effect of uncertainty leads to variability in the system performance or the process outcomes**
  - **Uncertainty quantification, through stochastic computational modeling and analysis, is an important tool for investigating these effects**

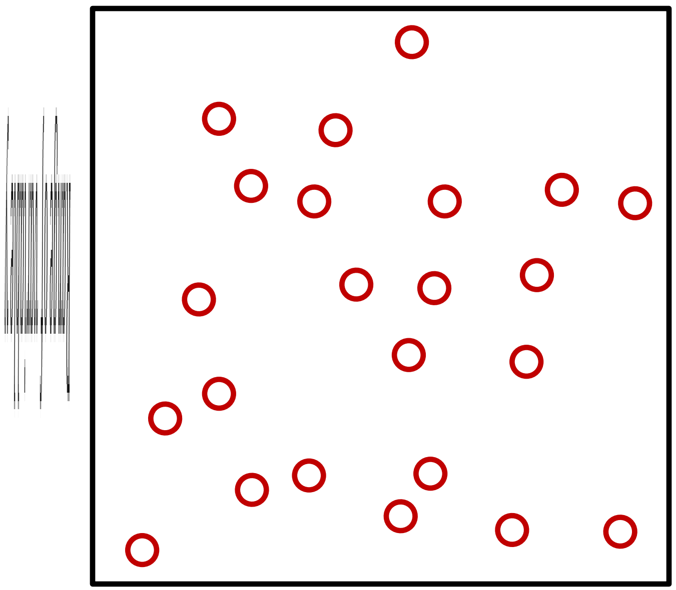


Current Methodologies

- Conventional methods seek to mimic physical processes
  - Monte Carlo methods randomly select inputs from the input distributions
  - Stratified methods (such as the Latin Hypercube method) seek to reduce the number of experiments, but still generate a representative sample
  - Both methods are VERY computationally intensive

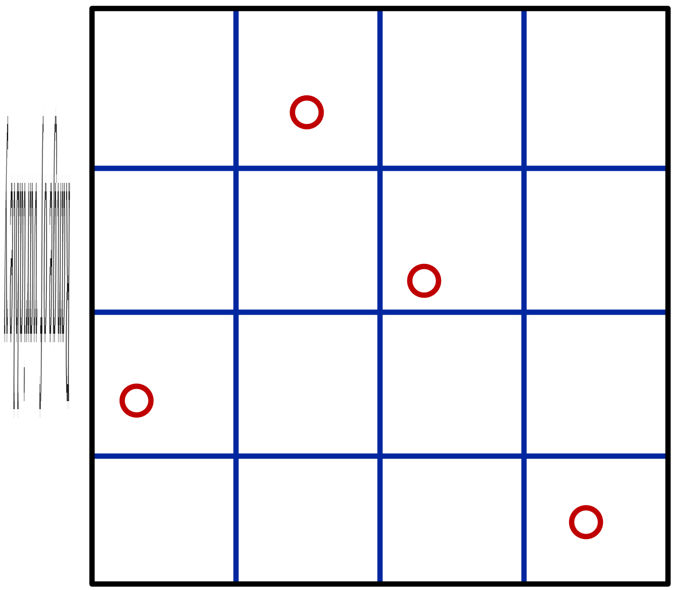
Monte Carlo Sampling

Range of variable x



Latin Hypercube Sampling

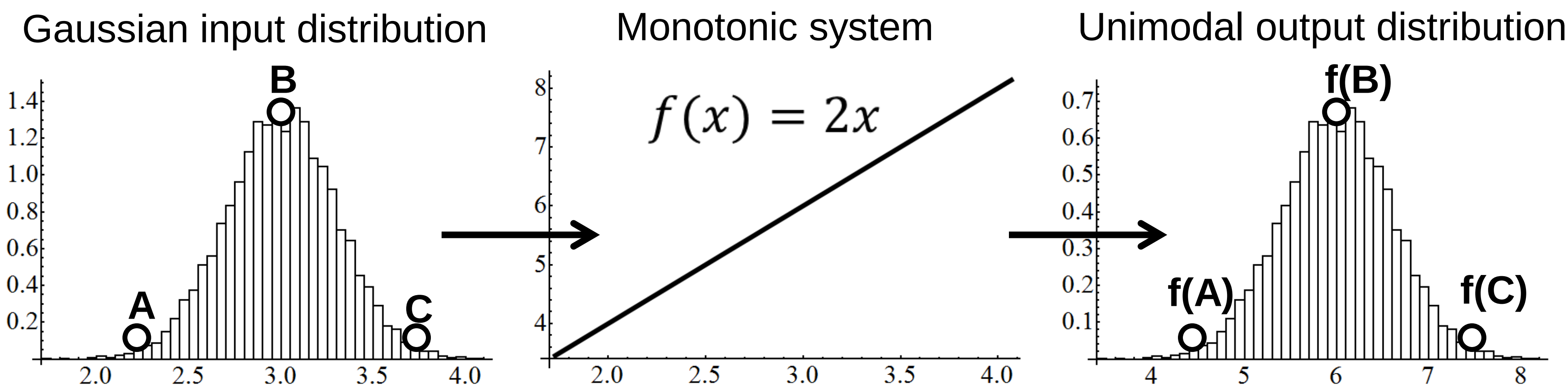
Range of variable x

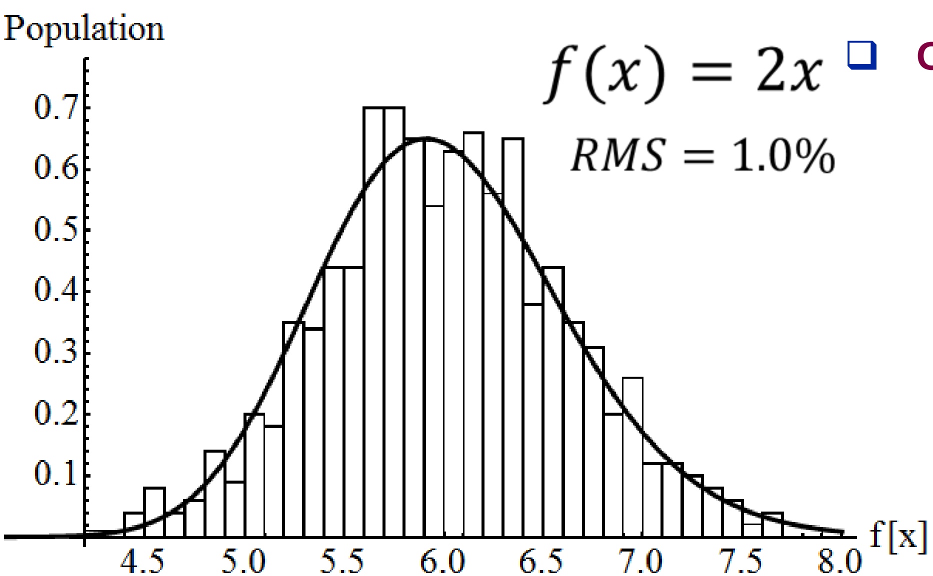


Overview of QUICKER

- QUICKER (Quantifying Uncertainty In Computational Knowledge Engineering Rapidly)
  - QUICKER is a new methodology that is intended to be used instead of conventional sampling methods such as Lathin Hypercube Sampling, Monte Carlo Sampling, Quasi-Monte Carlo Sampling, etc.
  - Since sampling, effectively running computational simulations, is the most time consuming aspect of Uncertainty Quantification, the significant reduction in computational costs from using QUICKER make Uncertainty Quantification far more affordable
- QUICKER is orders of magnitude faster than conventional sampling
  - Through the use of QUICKER, it is typical to see computational time reductions in excess of 99% of the time required for conventional methods
- QUICKER does not sacrifice accuracy
  - Typical RMS differences between QUICKER and conventional methodologies are less than 8%
- QUICKER is noninvasive and transparent
  - QUICKER can be implemented without modifying the simulation source code
  - The QUICKER methodology does not require esoteric math or complicated algorithms

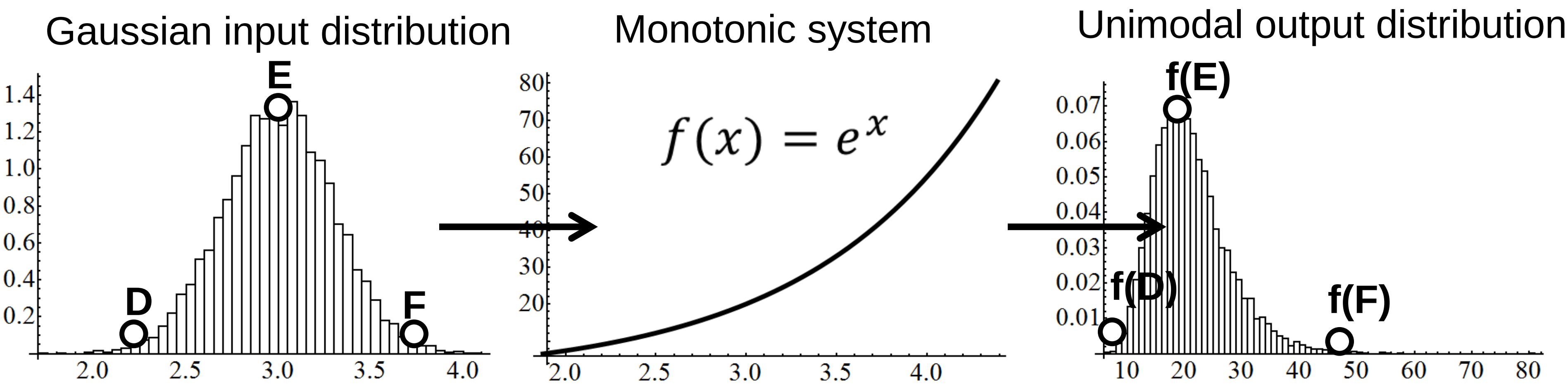
Identifying key points to sample

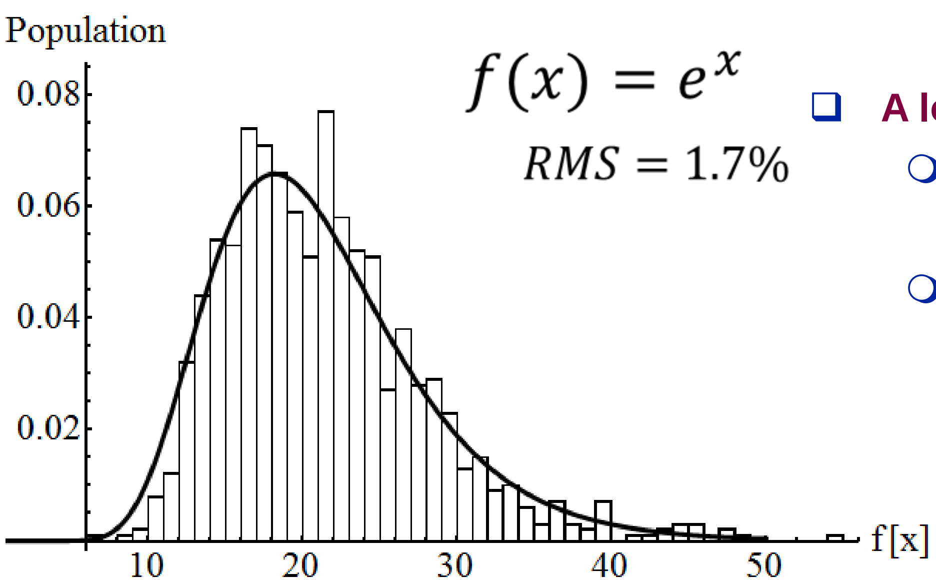




- **Only a small number of key points are necessary**
  - **For a monotonic system with a Gaussian input, it is necessary to select only a few input points in order to completely define the output distribution**
  - **These points are chosen at the mean and equal standard deviations**

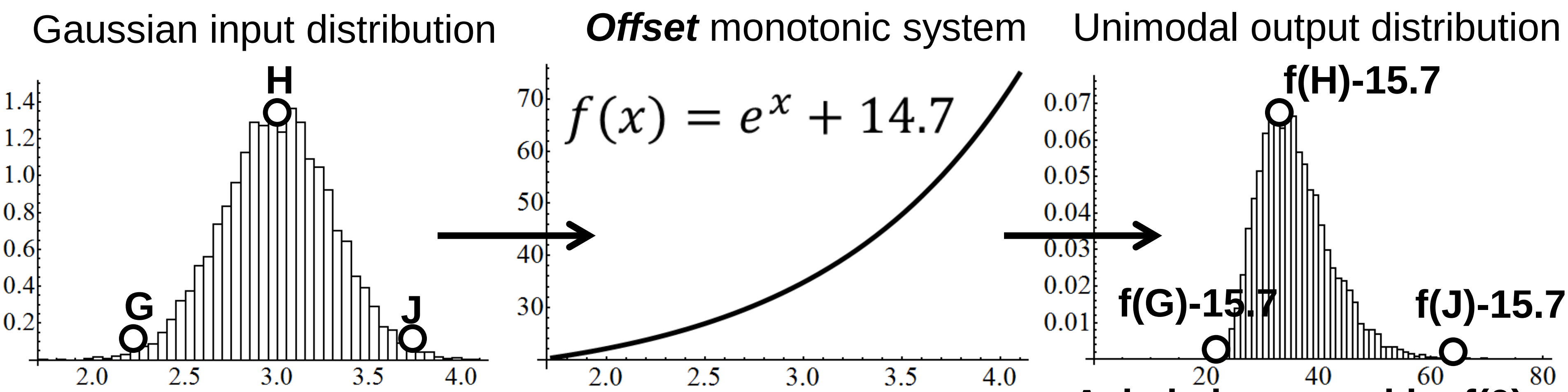
Using a lognormal output distribution



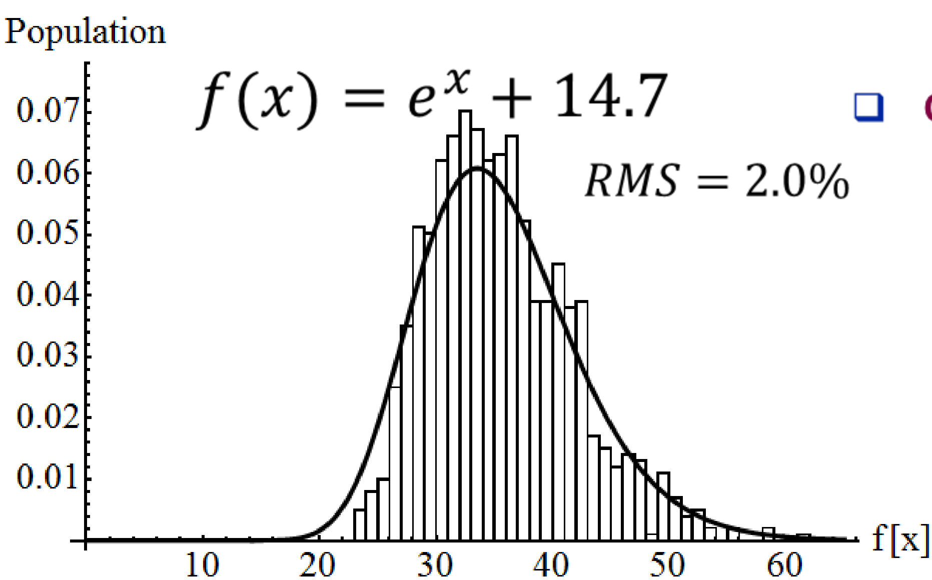


- **A lognormal distribution is versatile**
  - **A lognormal distribution can be used to represent symmetric or positive skewness**
  - **Therefore, lognormal distributions will be used in QUICKER**

Accounting for constant offset



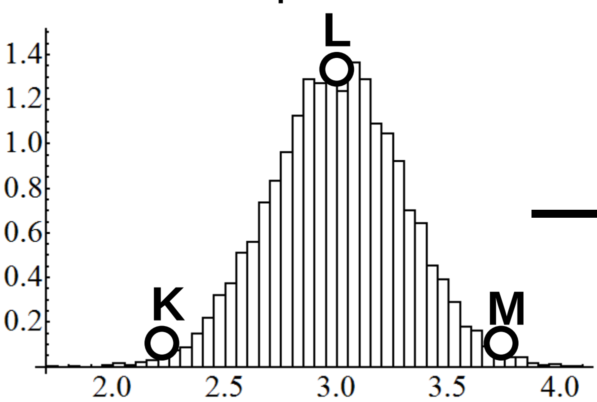
Axis is increased by f(0), 15.7 in this instance



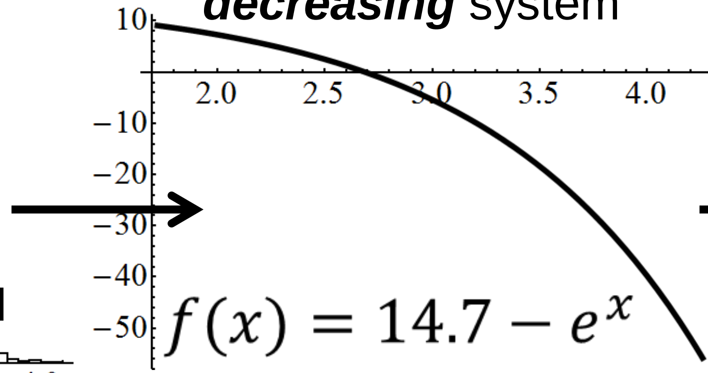
- **Certain systems have a constant offset**
  - **The lognormal distribution assumes that $f(x = 0) = 0$, and therefore it is necessary to account for any systematic offsets by taking an additional data point**

Accounting for negative skewness

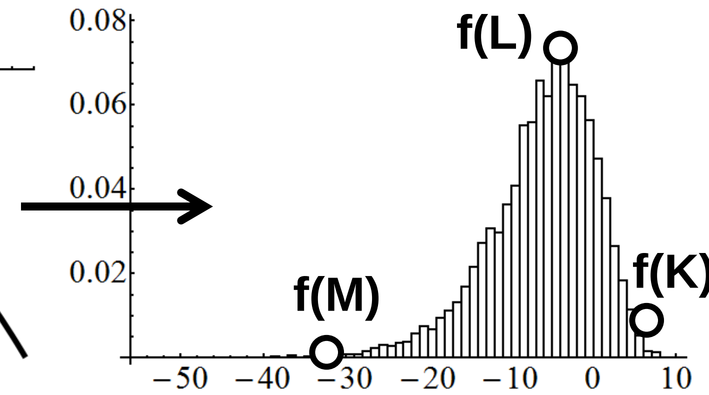
Gaussian input distribution



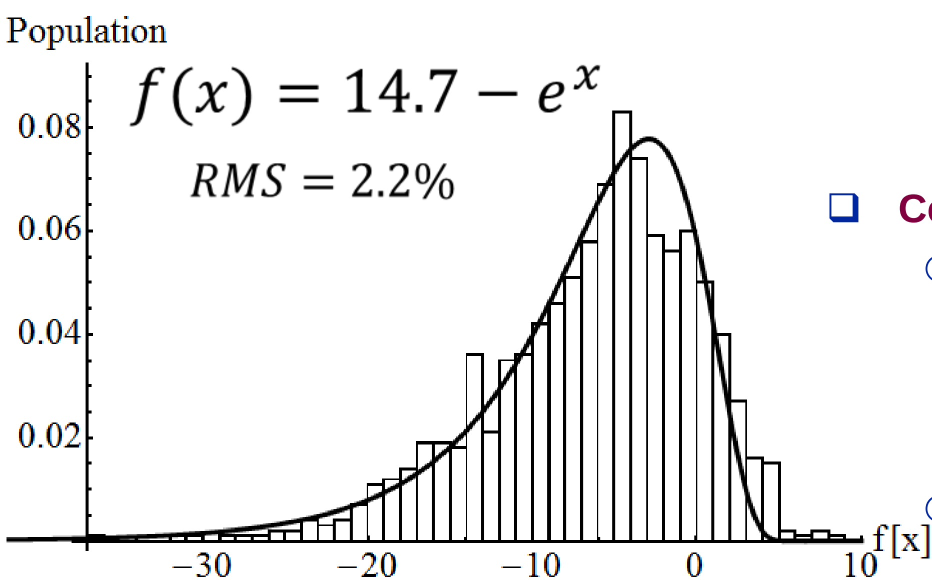
Monotonically *decreasing* system



Unimodal output distribution



Axis is flipped about the max result, f(K) in this instance

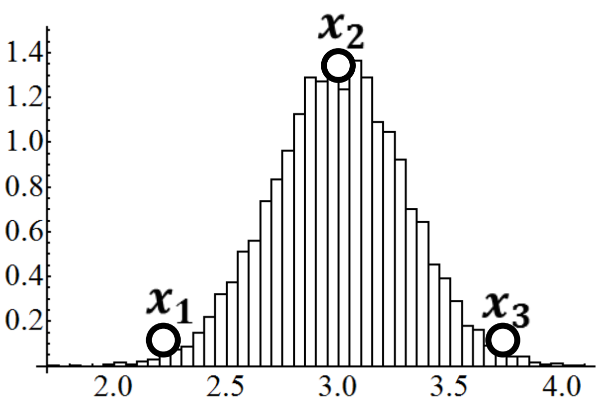


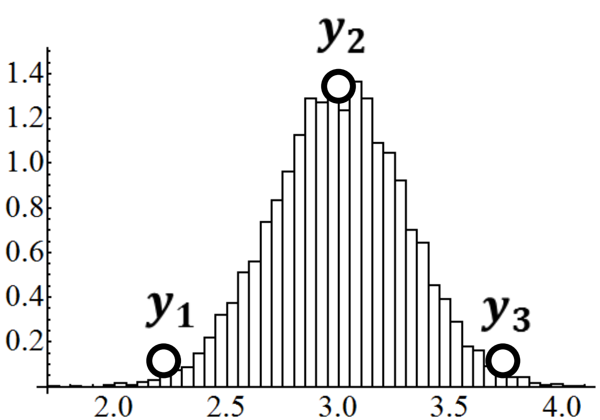
- **Certain outputs have a negative skewness**
  - **The lognormal distribution has a positive skewness, and in order to account for this, the plot needs to be “flipped” about the maximum point**
  - **Note that the function reverses the relative magnitude of the inputs**

Sampling within QUICKER

For a system with two input distributions, three points are selected on each input distribution

Input distr.	$\mu - N\sigma$	μ	$\mu + N\sigma$
x	x_1	x_2	x_3
y	y_1	y_2	y_3





The minimums and means are simulated, and then an orthogonal array is used to combine the extremes

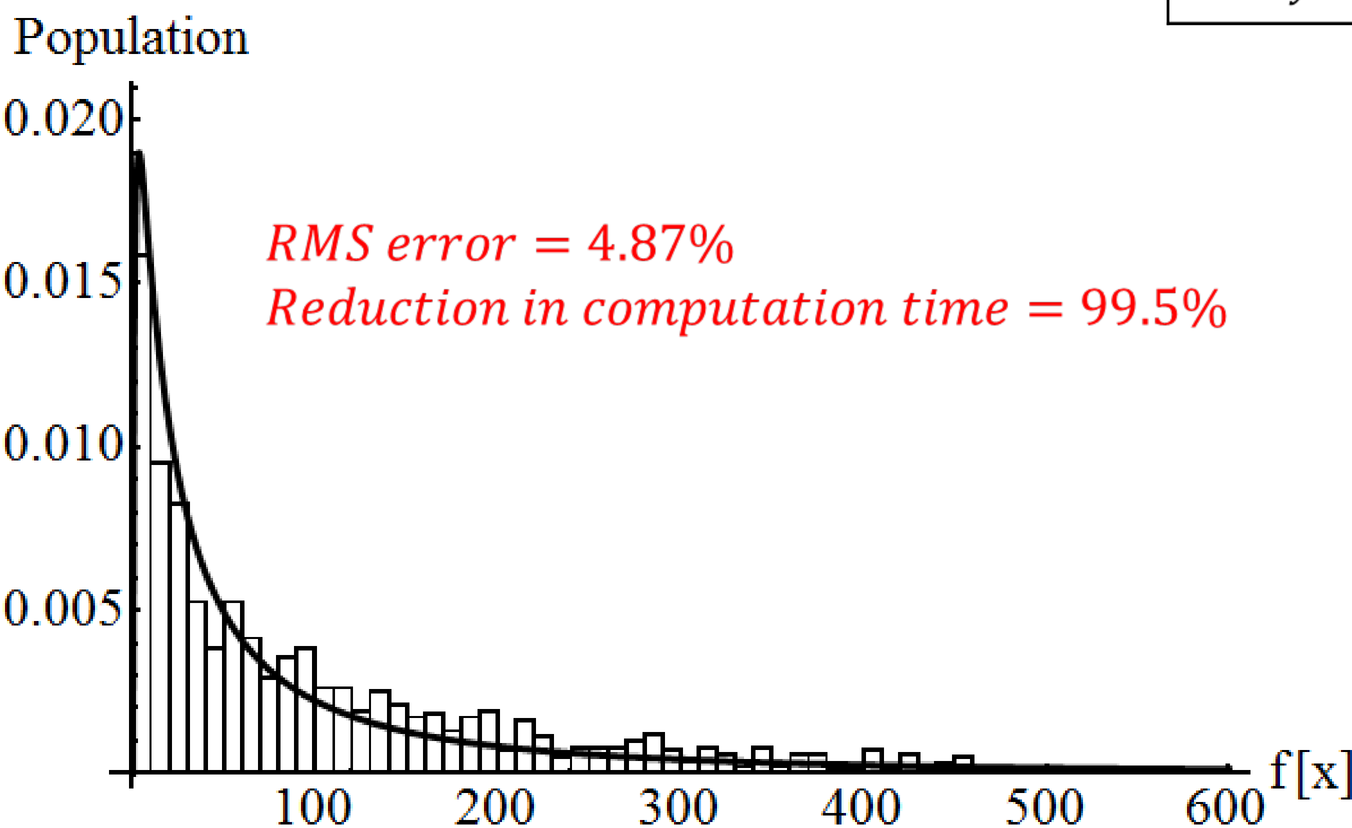
Simul. #	x	y
min	min	min
1	x_2	y_2
2	x_1	y_1
3	x_1	y_3
4	x_3	y_1
5	x_3	y_3

A specific example of QUICKER

The Rosenbrock function is a typical test case for optimization routines

$$f(x,y) = 100 * (y - x^2)^2 + (1-x)^2$$
$$\mu_x = 0,\ \sigma_x = 0.67$$
$$\mu_y = 1,\ \sigma_y = 0.67$$

Input distr.	$\mu - 3\sigma$	μ	$\mu + 3\sigma$
x	-2.01	0	2.01
y	-1.01	1	3.01



RMS error = 4.87%
Reduction in computation time = 99.5%

Simul. #	x	y
min	1	1
1	0	1
2	-2.01	-1.01
3	-2.01	3.01
4	2.01	-1.01
5	2.01	3.01

Composite representation of uniform

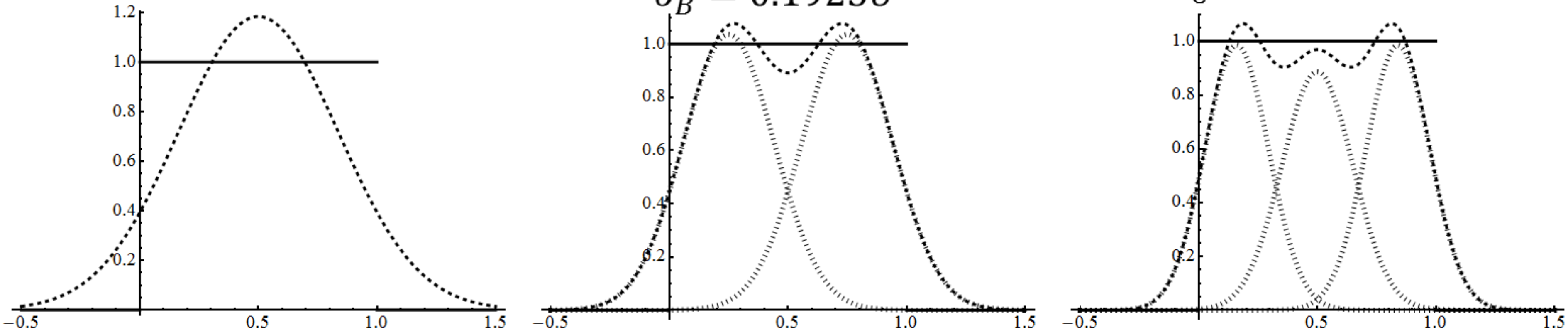
$U = length\ of\ uniform\ range$

$U_0 = lower\ bound$

$\mu = 0.5U + U_0$
$\sigma = 0.337U$

$\mu_A = 0.25U + U_0$
$\mu_B = 0.75U + U_0$
$\sigma_A = 0.1925U$
$\sigma_B = 0.1925U$

$\mu_A = 0.16U + U_0$
$\mu_B = 0.5U + U_0$
$\mu_C = 0.84U + U_0$
$\sigma_A = 0.135U$
$\sigma_B = 0.15U$
$\sigma_C = 0.135U$

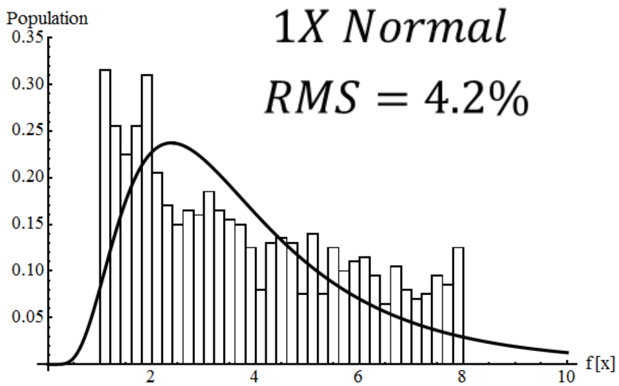


- Epistemic uncertainty is typically represented as a uniform distribution
  - A combination of Gaussian distributions can be used to represent a uniform distribution
  - Note that the Gaussian distribution is scaled depending on the uniform distribution

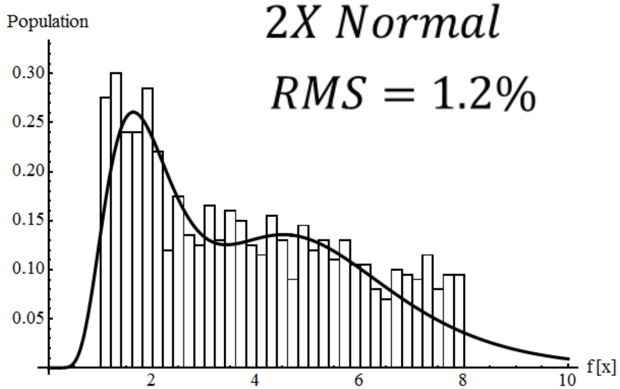
Results of composite uniform

$$Uniform\ input\ distribution: [1, 2] \qquad f(x) = x^3 + 1$$

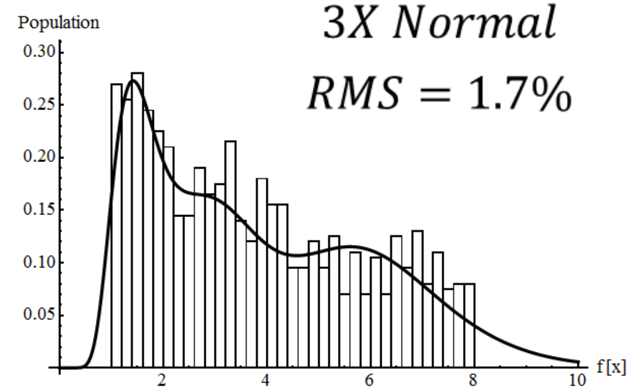
$Time\ reduction = 99.7\%$



$Time\ reduction = 99.4\%$



$Time\ reduction = 99.1\%$



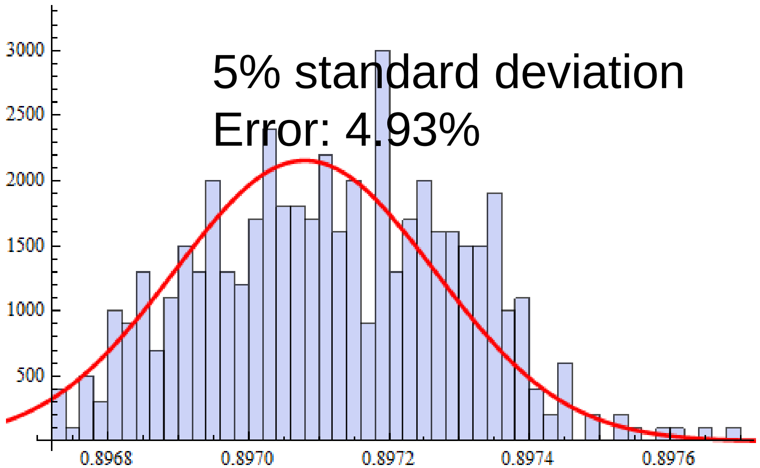
- **To a point, composite distributions can provide a more accurate prediction**
  - **Functions of only one variable are typically the hardest to represent with uniform composite distirbutions**
  - **Note that the improvement from 2X to 3X composite distribution is negligible**

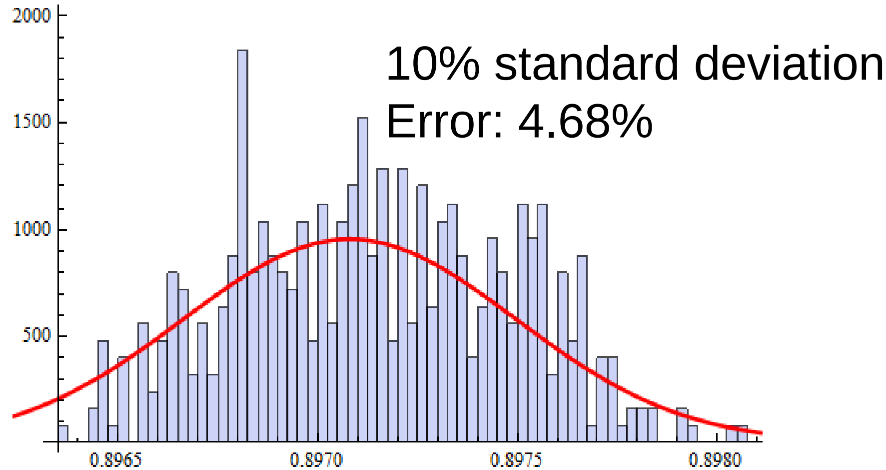
VirginiaTech
Invent the Future

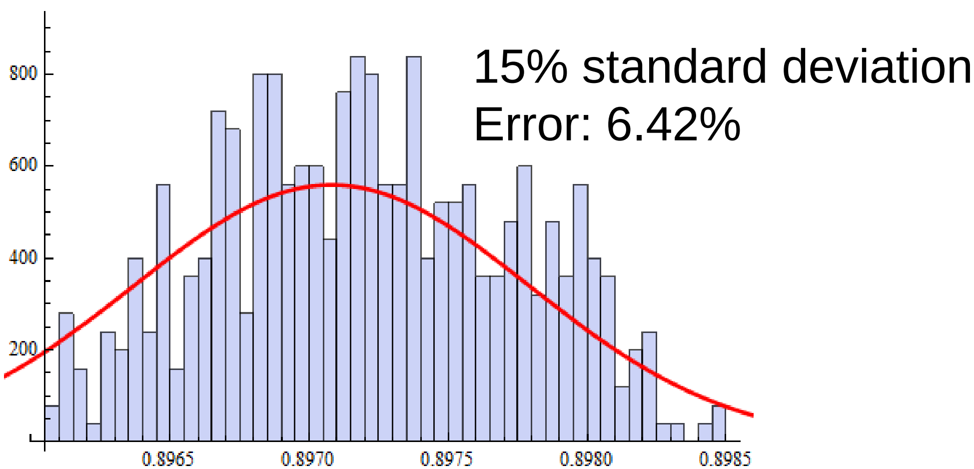
3dCfb MFIX scenario with 1X uniform

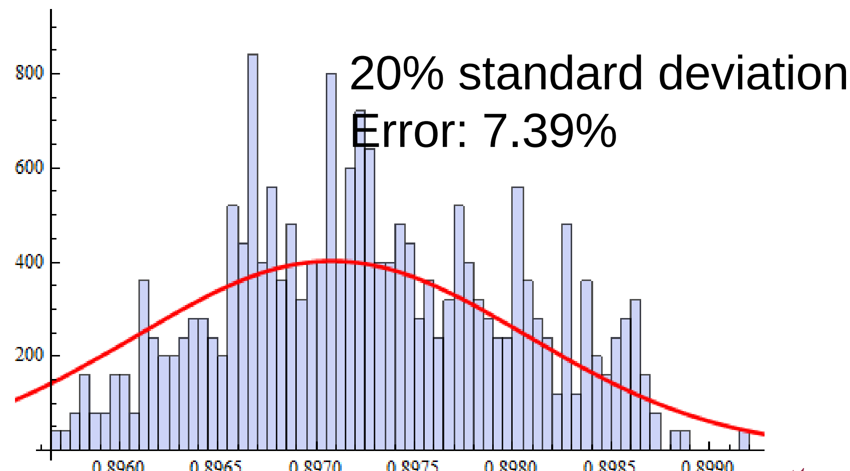
LHS computation time: 125.56 hours
QUICKER computation time: 4.52 hours
Computational time savings of *96.4%*

Measuring porosity at a specified location







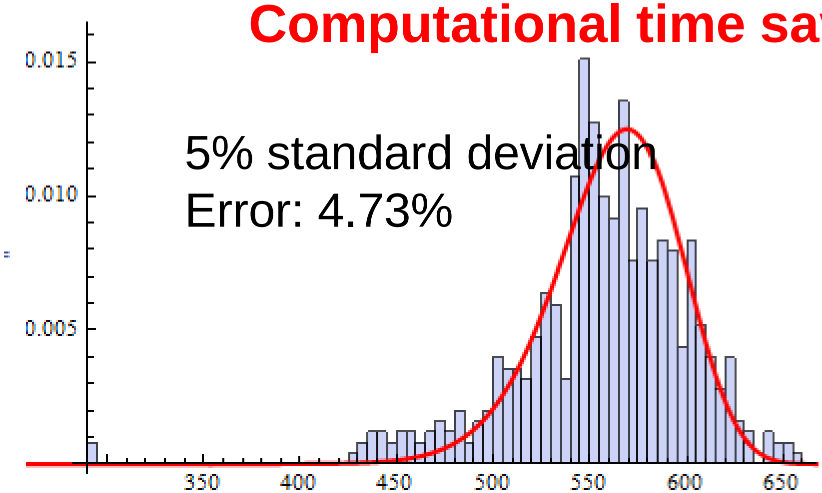


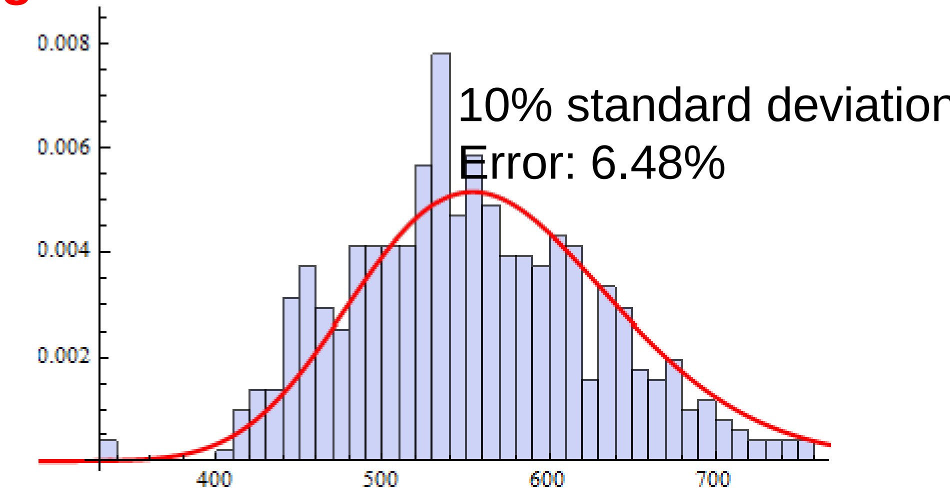
VirginiaTech
Invent the Future

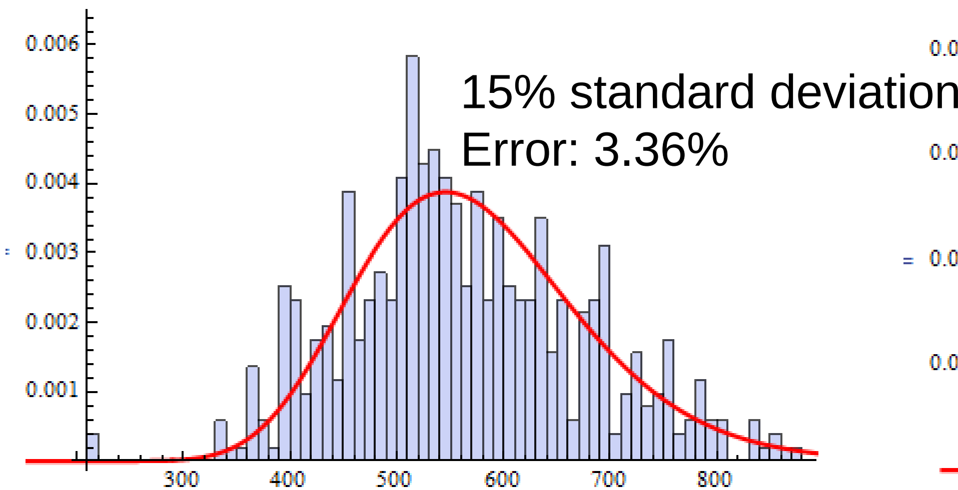
Ahmadi MFIX scenario with 1X uniform

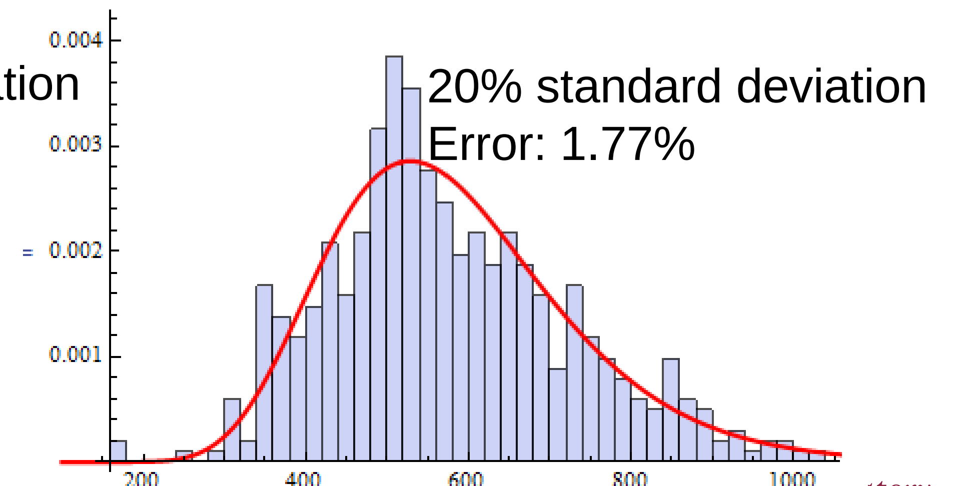
LHS computation time: 4.58 hours
QUICKER computation time: 0.10 hours
Computational time savings of 97.8%

Measuring solid velocity at a specified location







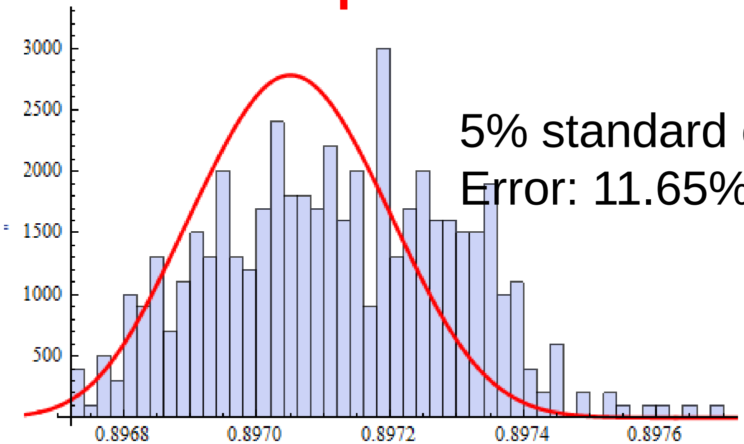


3dCfb MFIX scenario with 2X uniform

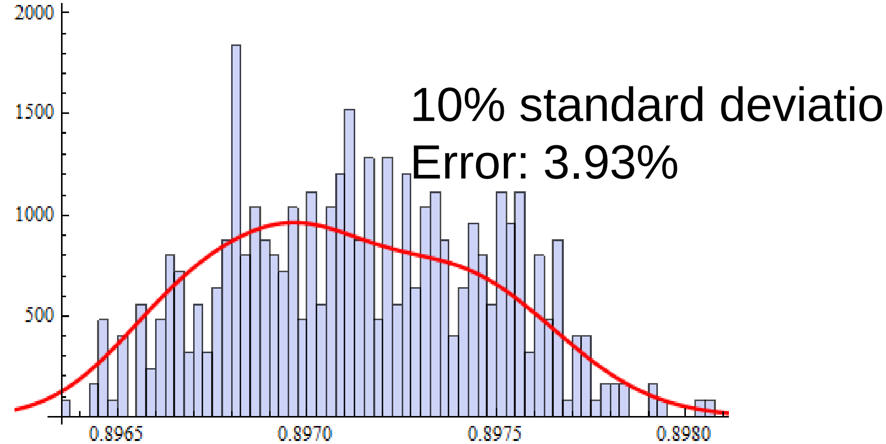
LHS computation time: 125.56 hours
QUICKER computation time: 36.16 hours
Computational time savings of 71.2%

Measuring porosity at a specified location

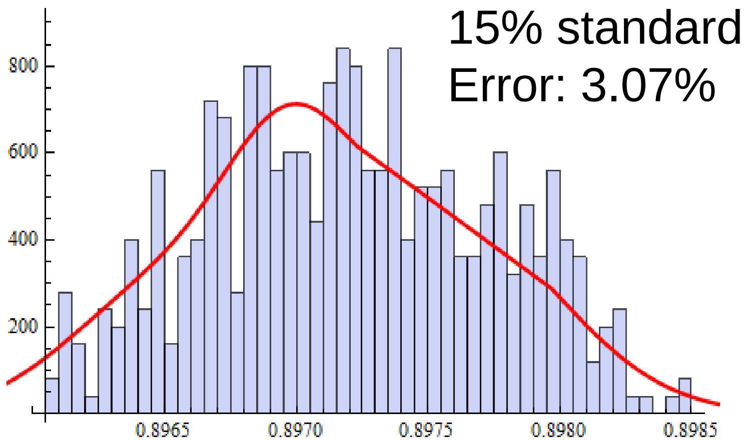
5% standard deviation
Error: 11.65%



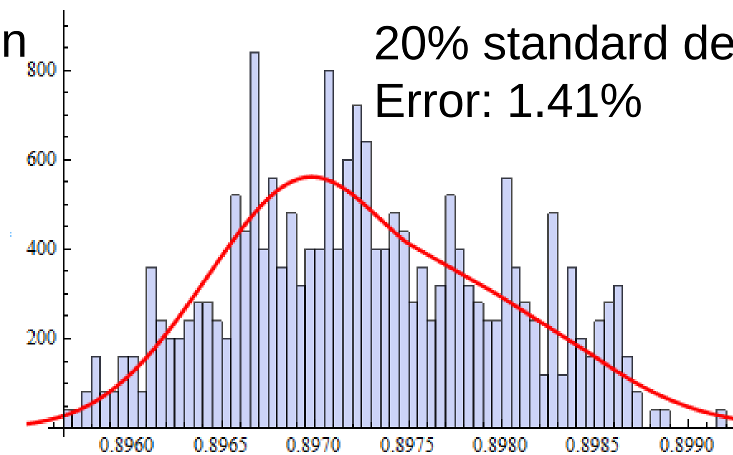
10% standard deviation
Error: 3.93%



15% standard deviation
Error: 3.07%



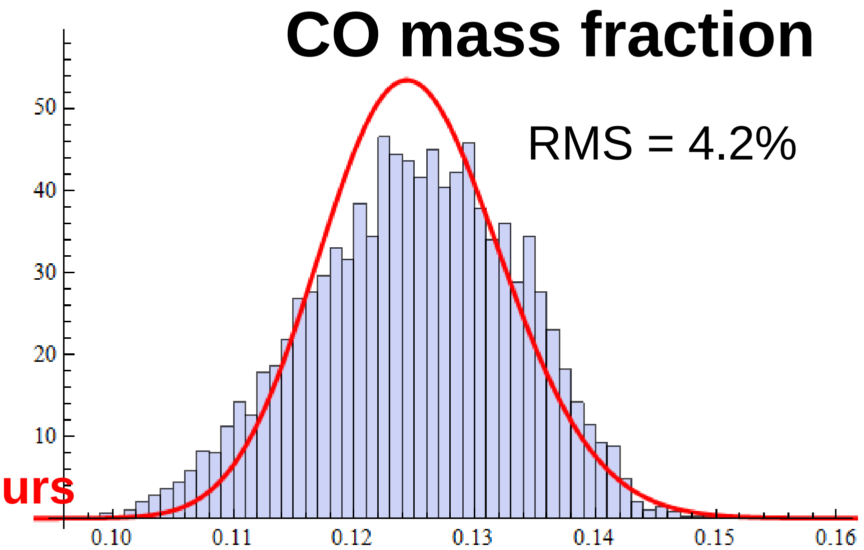
20% standard deviation
Error: 1.41%

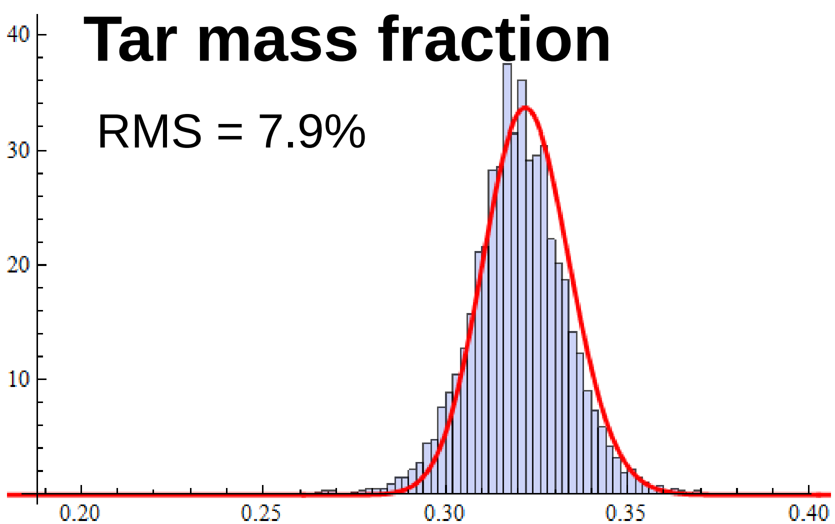


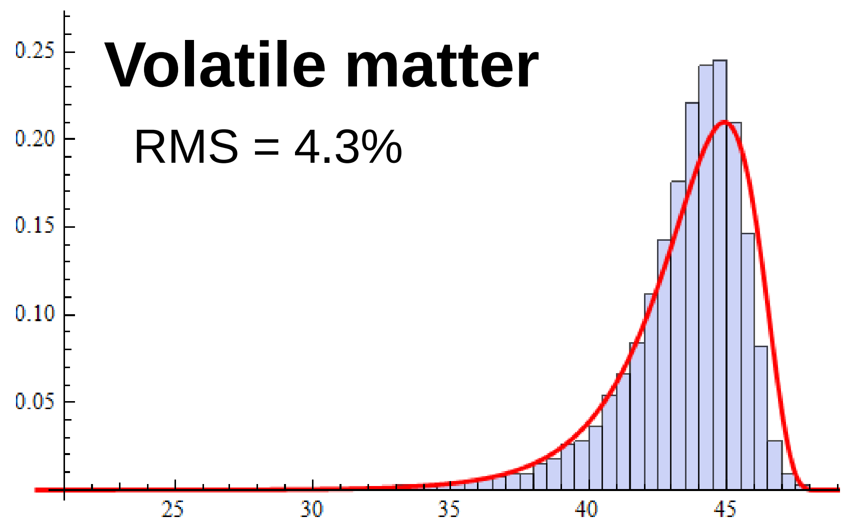
Results of a blind study

- **Results of a blind chemical kinetics study**
  - **Using the QUICKER methodology, a set of samples points were provided to Dr. Aytekin Gel to run through his simulation**
  - **The developers of QUICKER had no prior knowledge of the specifics of this kinetics model**

MC computation time: 1.52 hours
QUICKER computation time: 0.0019 hours
Computational time savings of 99.9%





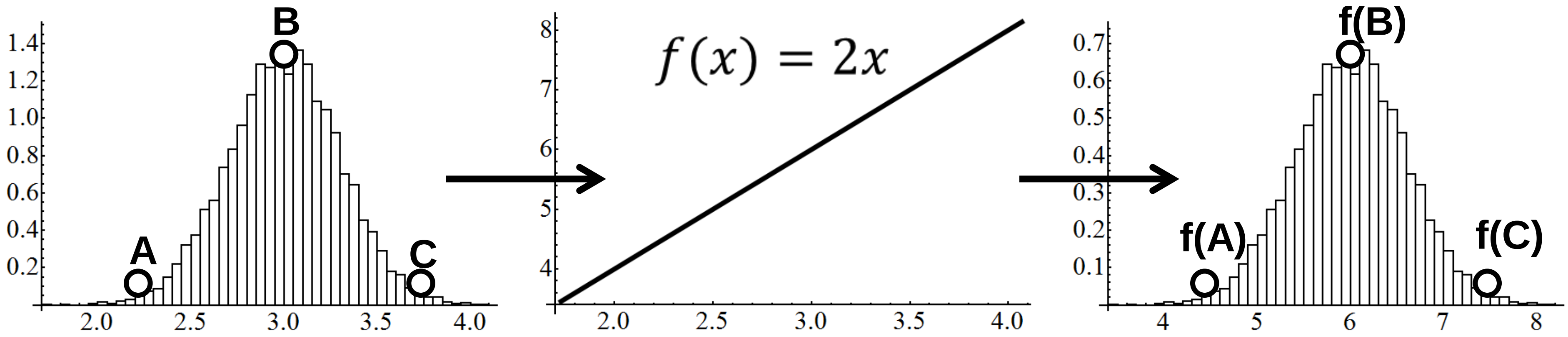


VirginiaTech
Invent the Future

Acknowledgements

Support from the U.S. Department of Energy, National Energy Technology Laboratory (NETL), is gratefully acknowledged

Gaussian input distribution | Monotonic system | Unimodal output distribution



Composite functions